# Archival and Data Distribution

How hackers and organizers can fight against online censorship & disinformation

An UnderGND.Online publication

# Introduction

If you're reading this zine, you're most likely aware of the rise of highly restrictive and privacy invasive Internet polices and legislation spreading all over the world. Multiple iterations of the Kids Online Safety Act (KOSA), which would require covered websites to "exercise reasonable care" to protect minors from certain harmful content and interactions, have been bouncing up and down the United States Congress, but half of US states have already passed age verification laws of some form.

The UK's Online Safety Act (OSA), a wide-reaching Act which imposes a "duty of care" on any "user-to-user service" to protect not only children but adults, was passed in 2023 and came into force in 2025. Australia's social media ban for under-16s took effect on 10 December 2025, and search engines will be required to verify the ages of their users in Australia to filter the results they display to users.

These laws introduce a threat to privacy as websites are required to collect identity documents or face scans of their users to satisfy their age verification obligations. But it goes further than that: UK's OSA contains provisions for the government's Internet regulator Ofcom to order scanning of

encrypted communications for illegal activity. At the time of writing (Dec 2025), the EU had just removed mandatory scanning requirements of users messages from their proposed “Chat Control” regulations, but they still contain provisions for voluntary scanning[1], laying the groundwork for mass surveillance.

Internet users have responded by using Virtual Private Networks (VPNs) to circumvent these restrictions, which has not escaped the notice of governments and legislators. The UK government is closely watching the use of VPNs, while the US states of Michigan and Wisconsin have introduced bills banning the use of VPNs in accessing “sexual content”[2].

A more recent and alarming development is the US Senate moving to repeal Section 230, a law which protects websites and users from being considered “publishers” when hosting or sharing other users’ content. The repeal would make such entities liable for content posted by others, effectively ending free speech on the Internet.

What explains this sudden rise of invasive and restrictive Internet policy? What can hackers and community organizers do, and how can information archival and distribution efforts counter the threats posed by these policies?

## Censorship & Book Bans

Let's be clear about what's really happening here: we are seeing a global, concerted effort across nation-states to enact censorship of the Internet.

These are the book burnings of the digital age. Consider this: who decides which kinds of content are considered harmful, objectionable, or not "age-appropriate"?

The term "age-appropriate" is often weaponized by far-right groups and legislators to restrict and censor the discussion of topics such as sexual orientation, gender identity, as well as reproductive and sexual health.

According to a report by PEN America[3], there were 6870 instances of book bans across 23 states and 87 public school districts in the 2024-2025 school year alone. The bans focused on books featuring people or topics involving marginalized populations such as people of color and LGBTQ+ people. Books dealing with sexuality and sexual violence were also targeted.

These bans are a way to exert top-down control over children and youths' access to information and ideas. It is also a way to erase certain populations such as queer and trans people from the public

sphere, beginning with youth exposure by labeling any presence or mention of LGBTQ+ people, identities or topics as inherently “sexual”, “obscene”, or even “pornographic”.

Targeting books about sexuality is a way to strip youths of their bodily autonomy and their ability to recognize and act against sexual and relational abuse and control perpetrated upon them by adults. One important domain of bodily autonomy is reproductive health and abortion rights, which has been under attack in the US since Roe v. Wade was overturned in 2022 and state-level abortion bans were passed or came into effect.

Amnesty International[4] has reported on how content on social media platforms containing information about abortions and abortion access has been removed, temporarily hidden, or marked as “sensitive content” by those platforms after the fall of Roe v. Wade.

These book bans have not stopped at schools: book-banning activists in the state of Utah have proposed that non-profit and community-run book exchanges and libraries such as Little Free Libraries should also be covered under the book bans[5], with the owners of those libraries made criminally liable for having banned books in their

libraries. It’s not a stretch to foresee this being extended to the digital domain.

Access to information is access to resources, and can be a matter of life and death. Folks need to know who or where to turn to when facing sexual abuse or violence, how to get abortion pills or access abortion across jurisdictions, or how to safely obtain and administer DIY hormone replacement therapy, among other things. What do you do when information itself becomes illegal to access, share, or even hold?

## Establishing ground truth

The rapid adoption of generative AI has given rise to a whole host of economic, environmental and municipal impacts, such as its enormous water usage, as well as causing power bills to soar in areas close to AI data centers. People also recognize that the probabilistic nature of generative AI has created a deluge of low-quality and inaccurate “AI slop” flooding the Internet.

Current AI-generated content displays certain characteristics that people believe can be used to identify it as AI-generated with a decent degree of confidence. For instance, AI-generated text may overuse specific words or have overly “mechanical” sentence structure in text, and AI-generated images

and video may have inconsistencies and errors like distorted lettering or deformed anatomical features, which a human would not make. However, generative AI models are constantly being refined, and these “tell-tale” signs may well be reduced or eliminated over time.

While much has been said about these issues, the inconsistent and inaccurate nature of AI-generated content also poses a deeper and less recognized threat: If you can’t tell whether a specific piece of content was AI generated, or you can’t verify the veracity of information being published, what does that do to our ability and methods to determine what is true or false?

In short: How does one establish a set of “ground truths” when considering any matter?

Low-quality and inaccurate AI generated material is already plaguing domains such as academia. At the 2024 PharmSci 360 conference hosted by the American Association of Pharmaceutical Scientists (AAPS), academics found that “AI is powering an arms race in the world of research misconduct”, and Special Issue Journals had been “overrun with articles submitted via paper mills, so much so that more than 8,000 papers had to be retracted”[6].

This can easily overwhelm the capacity of humans to rigorously review and validate academic content, and erode trust in academia. Trust in science has been a strong predictor of COVID-19 vaccine acceptance in the US[7], where it is estimated that hundreds of thousands of deaths during the COVID-19 pandemic could have been averted[8] if the least-vaccinated states had vaccinated at the same rate as the most-vaccinated states (52% vs 73%). If trust in science and academia is further eroded, the consequences can be destructive for society.

Another issue less often considered is the intentional usage of AI to generate disinformation. This can be employed by nation-state actors, as well as those who own and control our tech stacks, to influence not just other societies but even their own. This is a sense-breaking tool to distort people's ability to understand what is going on politically and economically; if the people do not understand what's happening around them, they can't organize and uplift themselves amid the systems that oppress and exploit them.

Archives thus serve a vital purpose in establishing a set of ground truths. They keep a historical record of web content that persists even if the content is later edited or removed, providing a bulwark against

historical revisionism by far-right governments and politicians. They can be used to maintain a repository of media and materials from before the generative AI boom that people can trust has not been adulterated by AI-generated content.

It is no wonder then that archival services are under attack. The Internet Archive was sued by a group of book publishers for copyright infringement in 2023[9], resulting in a settlement and removal of over 500,000 books from the archive. Record labels then sued the Internet Archive in 2024 for their 78rpm record archive[10], which could have bankrupted the Internet Archive but ended in a confidential settlement. The Internet Archive was later temporarily taken down by a DDoS attack in October 2024.

In November 2025, another archival website, archive.today, was reported to be the target of a subpoena and criminal investigation by the US Federal Bureau of Investigation (FBI)[11], demanding information from archive.today's domain registrar to identify the owners of the website.

## How can we build archives?

All of this underscores the crucial need for distributed, decentralized, and grassroots archival and data distribution efforts. These ideas are not

new, but it is crucial to increase their reach, accessibility, and urgency, particularly in the current context.

The traditional understanding of a digital archive is some sort of server farm, with multi-terabyte collections of books, software, academic papers, film, etc. which are set up and run by private individuals and made accessible over the internet. This may already exist in your communities of interest; if so, you may hopefully be motivated to join their efforts to lend a hand, or even help fund their efforts.

But archives do not necessarily need to look like multi-terabyte collections of media. For example, you might be inclined to run your own website distributing information about reproductive rights and abortion access.

Private individuals or communities can also run self-hosted file-sharing services, which can then be used to store and distribute various kinds of materials such as documentaries, zines, etc.

One way to set up your own file-sharing service is an open-source file-sharing and collaboration system called Nextcloud, which you can host on your own hardware. In fact, this very zine you're reading right now was written by multiple

collaborators over a self-hosted Nextcloud instance. A Nextcloud instance can be set up on a spare computer, or on hardware bought from a computer recycler.

Decentralized peer-to-peer file-sharing protocols and services such as BitTorrent and the InterPlanetary File System (IPFS) can also be used to distribute not only files, but websites as well. When Wikipedia was blocked in Turkey from 2017 to 2020[12], IPFS was used to create a mirror of the Turkish Wikipedia[13]. The mirror has now been expanded to include English, Russian, Ukrainian, Arabic and Chinese language versions of Wikipedia.

But what if it becomes unsafe to run such archives? What if access to archives or usage of specific protocols becomes blocked by technological means?

Deep Packet Inspection (DPI) technology has already been employed by several countries such as China, Russia, Singapore, and Indonesia in order to surveil, censor, and block internet traffic. The US National Security Agency has also been known to work with US internet service providers such as AT&T to employ DPI to conduct surveillance of internet traffic[14].

As we have previously highlighted, VPNs are already on the radar of legislators in the context of circumvention of age verification policies. It is not a stretch that technological and other measures might be employed in the future by various jurisdictions to block access to websites, archives, or even entire traffic protocols such as VPN and Tor.

This calls for a return to physical distribution of digital media, also known as “sneakernet”.

Physical distribution applies quite naturally to the physically smaller types of archives such as USB flash drives and portable hard drives. Pass them around your communities, or meet up at an agreed location, occasion or event to do a handover.

We can also bring back the idea of dead drops: a designated location containing a flash drive containing certain material, which can then be disclosed to those who need it. Such dead drops are usually inconspicuous and out of the way, like an abandoned utility box or a literal hole in the wall.

What about people or communities that maintain large archives? Well, physical distribution is arguably even more applicable. If you need to replicate a 20TB archive for instance, it’s going to take days even over a gigabit fiber connection. This is where magnetic tape comes in.

Older generation tape such as LTO-6 or LTO-7 has substantial native capacities of 2.5TB or 6TB, and up to 2.5x compressed capacities, while being reasonably priced. The tape can be passed around or even mailed long distance, for an archive to be replicated at a remote location. Incremental updates to the archive could then be feasibly synced on the wire. This will be important for decentralizing and distributing archives for resiliency against legal or physical takedowns.

Let's also not forget that all of this work we're doing is merely the medium to store and transmit what's truly important: information. While we are hackers of the digital realm, we should also make full use of analogue media to fulfill our ends. For instance, leaflets can be hidden in books and magazines in stores or libraries, or coded/ciphered messages can be hidden in paper letters. This can be used to disseminate short forms of vital life-saving information, or point readers to larger archives/repositories/communities.

This raises the question of how to safely conduct such activities or safely communicate and decide on physical locations and meetings. This is not a technical question, but a social one, which will require social solutions. How do you build communities and networks of trust, and how do you

scale them up to reach those whose needs would be met by such communities and networks?

The answers will vary by context, and are beyond the scope of this zine, but we will emphasize the importance of operational and information security, vetting and onboarding processes, segmented levels of access and trust, accountability, and mutual care in cultivating safe and resilient communities and networks.

## Conclusion

The attacks on privacy, freedom of information and epistemic stability are happening in real time and must be taken seriously. At the same time, they also don't happen overnight. We must make the most of the current window of free access to information and private communication to retain and archive the information that is important to us, and more importantly, form the communities and networks, both local and remote, both physical and virtual, so that access to information and resources can be established, sustained, and renewed into the future.

In short: ORGANIZE!

# References


[1] https://www.eff.org/deeplinks/2025/12/after-years-controversy-eus-chat-control-nears-its-final-hurdle-what-know

[2] https://www.eff.org/deeplinks/2025/11/lawmakers-want-ban-vpns-and-they-have-no-idea-what-theyre-doing

[3] https://abcnews.go.com/US/report-warns-disturbing-normalization-book-bans-us-schools/story?id=126152594

[4] https://www.amnesty.org/en/latest/news/2024/06/united-states-social-media-companies-removal-of-abortion-related-content-may-hinder-access-to-accurate-health-information/

[5] https://www.themarysue.com/utah-book-banners-now-want-to-make-little-free-libraries-susceptible-to-criminal-charges/

[6]https://www.aapsnewsmagazine.org/aapsnewsmagazine/articles/new-page3/feb25/meetings-feb25

[7] https://www.sciencedirect.com/science/article/pii/S2590136224001499

[8] https://www.cidrap.umn.edu/covid-19/us-pandemic-death-toll-higher-20-peer-countries

[9] https://en.wikipedia.org/wiki/Hachette_v._Internet_Archive

[10] https://www.rollingstone.com/music/music-features/internet-archive-major-label-music-lawsuit-1235105273/

[11] https://www.404media.co/fbi-tries-to-unmask-owner-of-infamous-archive-is-site/

[12] https://en.wikipedia.org/wiki/Block_of_Wikipedia_in_Turkey

[13] https://blog.ipfs.tech/24-uncensorable-wikipedia/

[14] https://en.wikipedia.org/wiki/Room_641A